Daubert, Critique and Interpretation: What Empirical Studies Tell Us About the Application of *Daubert*

By A. LEAH VICKERS*

LEGAL SCHOLARS HAVE had a field day with the Supreme Court's 1993 decision in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*[1] This opinion, which changed the standard for the admissibility of expert evidence in federal courts and the judge's role in admissibility determinations,[2] has inspired reams of literature and been subject to a variety of criticisms. Most of these criticisms predict that *Daubert* would result either in the excessive exclusion of evidence, or in arbitrary or inconsistent distinctions between what evidence is, and is not, admissible. The majority of these criticisms address the context of torts, particularly toxic torts, and this Article reflects this focus.

Until 2001, these criticisms of *Daubert* necessarily occurred only at a theoretical level and at best with anecdotal evidence. Thankfully, in the past few years, at least three comprehensive empirical studies have shed light on how judges and parties to litigations are responding to the decision.[3] A fourth study by Professors Edward K. Cheng and Albert H. Yoon provides critical and surprising links between the data developed in prior studies.[4] Taken together, these studies show that

* Associate (admission pending), Sonnenschein Nath & Rosenthal LLP, New York; B.A., Columbia University 2000, *magna cum laude*; J.D., Stanford Law School 2005. I would like to thank Professors Barbara Fried and Tom Grey, my classmates in the Legal Studies Colloquium, and the editors of the U.S.F. Law Review for their valuable input. I would also like to thank my family and my life-long editor, Kirsten Olds.

1. 509 U.S. 579 (1993).

2. *Id.* at 597–98.

3. Lloyd Dixon & Brian Gill, *Changes in the Standards for Admitting Expert Evidence in Federal Civil Cases Since the* Daubert *Decision*, RAND Institute for Civil Justice Report (2001) [hereinafter RAND Study]; Jennifer L. Groscup et al., *The Effects of* Daubert *on the Admissibility of Expert Testimony in State and Federal Criminal Cases*, 8 PSYCHOL. PUB. POL'Y & L. 339 (2002) [hereinafter Groscup Study]; Carol Krafka et al., *Judge and Attorney Experiences, Practices, and Concerns Regarding Expert Testimony in Federal Civil Trials*, 8 PSYCHOL. PUB. POL'Y & L. 309 (2002) [hereinafter FJC Study].

4. Edward K. Cheng & Albert H. Yoon, *Does* Frye *or* Daubert *Matter? A Study of Scientific Admissibility Standards*, 91 VA. L. REV. 471 (2005), [hereinafter Cheng & Yoon Study].

after *Daubert*, parties challenged the admissibility of evidence more frequently, and judges scrutinized evidence more carefully, excluding a greater proportion of it.[5] While none of these studies answers the looming question of whether judges are excluding too much evidence, this data does provide a greater understanding of the mechanisms by which *Daubert*, as applied by judges, has heightened the bar for admissibility and, accordingly, reduced the proportion of evidence deemed admissible. A careful consideration of this information about how *Daubert* is actually applied provides insight into which, if any, criticisms of *Daubert* correctly identify the doctrine's purported weaknesses.[6] *Daubert* has had a profound effect on the admissibility of evidence but not via the means that most critics would guess. In fact, *Daubert*'s impacts appear to be the result not of the doctrinal test set forth in the decision, but rather of a cultural phenomenon either sparked by the decision, or to which the decision has contributed. Thus, criticisms aimed at the text itself, specifically at the criteria for reliability (which are not frequently utilized), would seem to miss the mark. The better questions to ask are why the decision failed to provide criteria that judges actually find useful in drawing these distinctions and what criteria judges are, in fact, using to distinguish between admissible and inadmissible evidence.

Part I of this Article explains the progression of admissibility standards for expert testimony. Part II provides a general survey of common criticisms of *Daubert*, particularly in the toxic tort context, such as problems with the criteria for reliability set forth in the decision, *Daubert*'s excessive allocation of power to judges, and the judicial exclusion of otherwise valid data. Despite the great volume of paper produced by this discussion, no article, to this author's knowledge, has offered a comprehensive and conceptually organized overview of such criticisms (some of which cannot be empirically tested).[7] Part III proceeds by reviewing the relevant findings of recent empirical studies of *Daubert*'s effects and discusses the contribution these findings make to

5. *See generally* RAND Study, *supra* note 3; FJC Study, *supra* note 3; Cheng & Yoon Study, *supra* note 4.

6. Unfortunately, some of these criticisms are too nuanced to be addressed by the relatively general data produced thus far, and scholars concerned with the impacts of *Daubert* should be looking to conduct more detailed studies, particularly in the area of toxic tort litigation.

7. While it may seem one-sided to look only at the criticisms of *Daubert* without explicitly considering the benefits of this doctrine (particularly when compared to the alternatives), this Article operates under the assumption that since any doctrine is flawed in certain ways, the discourse surrounding *Daubert* may be aided by a careful summary of its alleged failings and consideration of whether these are borne out by empirical research.

our understanding of how *Daubert* affects admissibility. Part IV then considers how this data relates to criticisms of *Daubert*. Specifically, it examines the results of a recent study by Professors Cheng and Yoon and argues that while its methodology is novel, the study's results affirm the hypothesis that *Daubert*'s profound impacts have resulted less from *Daubert*'s actual doctrinal test and more from the fact that the decision increased awareness of, and concern about, junk science. This observation goes a long way in explaining how a decision that was meant to reflect the liberalized rules of evidence has, in fact, raised the bar to admissibility. It also begs the serious question of why the decision did not create a more useful doctrinal test. Before concluding, this Article also briefly makes recommendations for future research.

I. The *Daubert* Decision

In 1993, the Supreme Court issued its decision in *Daubert*,[8] dramatically reshaping the relationship between experts and the courts, and between law and science.[9] Prior to *Daubert*, the most commonly applied standard for admissibility of expert evidence derived from *Frye v. United States*.[10] *Frye* held that the proper inquiry is whether the scientific principle or technique from which the evidence is deduced is generally accepted in the particular field from which it arises.[11] The *Frye* test calls for "judges to decide the admissibility of scientific expert testimony by deferring to the opinions of scientists in the 'particular field.'"[12] It does not require that judges "have any facility with scientific methods to make the admissibility decision," but only mandates that they develop "some basis for knowing what [it is that] most scientists believe."[13]

Following the adoption of the Federal Rules of Evidence ("Federal Rules") in the 1970s, courts began questioning the continued applicability of *Frye*.[14] Simultaneously, critics of the *Frye* test alternately accused it of being too conservative—requiring that courts wait until a

8. 509 U.S. 579 (1993).

9. Joseph Sanders et al., *Legal Perceptions of Science and Expert Knowledge*, 8 PSYCHOL. PUB. POL'Y & L. 139, 142 (2002) [hereinafter Sanders, *Legal Perceptions*].

10. 293 F. 1013 (D.C. Cir. 1923).

11. *Id.* at 1014; Sanders, *Legal Perceptions*, *supra* note 9, at 140.

12. 1 DAVID L. FAIGMAN, ET AL., MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY § 1-3.0, at 13 (2d ed. 2001) [hereinafter MODERN SCIENTIFIC EVIDENCE].

13. *Id.*

14. Sanders, *Legal Perceptions*, *supra* note 9, at 141.

method is generally accepted—or too lax—deferring to experts in fields that lack a tradition of rigorous scrutiny.[15] Against this backdrop, the Supreme Court granted certiorari to review the question of whether the Federal Rules, particularly Rule 702,[16] incorporated or displaced the *Frye* test.[17] Finding the "rigid 'general acceptance' requirement . . . at odds with the 'liberal thrust' of the Federal Rules and their 'general approach of relaxing the traditional barriers to 'opinion' testimony,'" the Court provided an interpretation of Rule 702 that effectively set a new standard for the admissibility of expert evidence.[18]

Specifically, *Daubert* requires that for evidence to be admitted it must be relevant ("assist the trier of fact to understand the evidence or to determine a fact in issue")[19] and must meet a standard of "evidentiary reliability."[20] In describing the reliability requirement the Court stated that the judge should conduct a "preliminary assessment of whether that reasoning or methodology properly underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue."[21] Although the Court acknowledged that this assessment will involve many factors and explicitly denied creating a checklist, it set out five specific factors that judges may want to consider: (1) whether the evidence is "falsifiable" and has been tested, (2) "whether the theory or technique has been subjected to peer review and publication," (3) what the known or potential rate of error of the technique is, (4) whether standards exist and are maintained for controlling this technique's operation, and (5) whether the methods and reasoning are generally accepted.[22]

15. 1 MODERN SCIENTIFIC EVIDENCE, *supra* note 12, § 1-3.0, at 13. In 1991, Peter Huber, a senior fellow of the Manhattan Institute for Policy Research, published a widely distributed book, GALILEO'S REVENGE: JUNK SCIENCE IN THE COURTROOM, accusing expert witnesses of introducing unsubstantiated claims into the courtroom and accusing the court system of failing to differentiate between these and valid evidence. *See generally* PETER HUBER, GALILEO'S REVENGE: JUNK SCIENCE IN THE COURTROOM (Basic Books 1991).

16. FED. R. EVID. 702. At the time, the rule stated: "If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise . . . ." *Id.*

17. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 588–89 (1993).

18. *Id.* at 588 (citing *Beech Aircraft Corp. v. Rainey*, 488 U.S. 153, 169 (1988)).

19. *Id.* at 591 (citing FED. R. EVID. 702). The relevance test had been the alternative to *Frye*. RAND Study, *supra* note 3, at 1.

20. *Daubert*, 509 U.S. at 590.

21. *Id.* at 592–93.

22. *See id.* at 593–95 (interpreted in Sanders, *Legal Perceptions*, *supra* note 9, at 141). Some commentators do not include criteria number 4 above—whether standards exist and are maintained for controlling this technique's operations—and hence only refer to four

Thus, *Daubert* moved judges into the role of gatekeeper, charged with the "responsibility of evaluating the scientific validity of the basis for expert testimony," and "obligated to become familiar with the methods and culture of science."[23] While under *Frye* judges sometimes held hearings to ensure that the proffered evidence met the general acceptance test, they essentially acquiesced to the judgment of the scientific community.[24] The decision effectively told them to "do whatever the experts tell you to do."[25] Today, under *Daubert*, judges are responsible for questioning the very methods and procedures of scientists. As Professor Sanders and others have written, the *Daubert* test allows judges to play a more active role and provides for a more nuanced analysis than is possible under *Frye*, but "it is also true that this very flexibility makes *Daubert* a more uncertain test."[26] This change has sparked significant debates among judges, practicing attorneys, legal scholars, and scientists over the past decade.

Before embarking on a summary of the criticisms that have been leveled at the *Daubert* decision, it is necessary to briefly mention two cases that followed *Daubert*: *General Electric Co. v. Joiner*[27] and *Khumo Tire Co. v. Carmichael*.[28] In these decisions, the Supreme Court further shaped and clarified the *Daubert* doctrine. Along with the original decision, these cases comprise what is often called the "*Daubert* Trilogy."

In *Joiner*, decided in 1997, the Supreme Court clarified that the standard of review for admissibility decisions (in other words, the standard appellate courts should utilize in reviewing trial court decisions) is "abuse of discretion." This is a less stringent standard than some courts had been applying.[29] Subsequently, in 1999, in *Kumho Tire*, the Supreme Court stated that "*Daubert*'s general holding—setting forth the trial judge's general 'gatekeeping' obligation—applies not only to testimony based on 'scientific' knowledge, but also to testimony based on 'technical' and 'other specialized' knowledge."[30] Broadly speaking, these decisions thus enhanced judges' power to apply *Daubert* by ex-

criteria for reliability. For the sake of consistency, I will refer to them as the five criteria or factors, though numerous quotes within this Article refer to them as "four" factors.

23. 1 MODERN SCIENTIFIC EVIDENCE, *supra* note 12, § 1–3.0, at 13.

24. Sanders, *Legal Perceptions*, *supra* note 9, at 142.

25. *Id.* at 143.

26. *Id.* at 142.

27. 522 U.S. 136 (1997).

28. 526 U.S. 137 (1999).

29. *See* Sanders, *Legal Perceptions*, *supra* note 9, at 143.

30. *Id.* at 141.

panding its application (or at least clarifying the breadth of its scope) and by helping to fortify trial courts' *Daubert*-based decisions.

II. Current Criticisms of *Daubert*[31]

A. *Daubert* Invokes Judicial Decision Making, Procedural, and Seventh Amendment Concerns

The first category of criticisms stems from concerns that *Daubert* has shifted too much power into the hands of judges. Perhaps the most common complaint about *Daubert* is that it forces judges to become amateur scientists, a role they are not well-prepared for and should not be asked to play.[32] In requiring a searching assessment of the methodology underlying the evidence, *Daubert* asks judges to acquire a much greater understanding of the science behind the evidence than was asked of judges and juries under *Frye*. Many, including the dissenters in *Daubert*,[33] have wondered whether it is wise to ask judges to undertake such an assessment when they have little training in scientific methods, reasoning, and research.[34]

To the extent that either the judge or jury must undertake some analysis of the relevance and reliability of proffered testimony, critics charge that by placing judges in the position of gatekeepers, *Daubert* unnecessarily usurps power from the jury.[35] This argument is generally framed as a question of decision making competency, but also contains constitutional undertones arising from the Seventh Amendment preservation of the jury trial.[36] Critics ask whether judges are any more competent than juries in assessing the reliability and relevance of scientific evidence. Finding that they may not be, they question

31. Although I have tried to provide some order to these criticisms by organizing them conceptually, there are significant overlaps, as reflected in the fact that discussions of certain lines of criticism are included in more than one section.

32. *See* Michel F. Baumeister & Dorothea M. Capone, *Admissibility Standards as Politics—The Imperial Gate Closers Arrive!!!*, 33 SETON HALL L. REV. 1025, 1039–40 (2003).

33. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 600–01 (1993) (Rehnquist, J., dissenting).

34. Certain categorical judgments regarding admissibility by judges in recent years have been criticized by scholars knowledgeable about scientific research. *See infra* Part II.B.

35. David M. Malone & Paul J. Zwier, *Epistemology After* Daubert, Khumo Tire, *and the New Federal Rule of Evidence 702*, 74 TEMPLE L. REV. 103, 106 (2001).

36. TELLUS INST., DAUBERT: THE MOST INFLUENTIAL SUPREME COURT RULING YOU'VE NEVER HEARD OF, A PUBLICATION ON SCIENTIFIC KNOWLEDGE AND PUBLIC POLICY 8 (2003) [hereinafter TELLUS PUBLICATION].

Daubert's wisdom of effectively taking this power away from the jury, particularly in light of the Seventh Amendment.[37]

Scholars assert that for epistemological reasons it may be preferable to permit the jury to review all minimally acceptable evidence together.[38] Some take issue with the post-*Daubert* practice of screening individual pieces of scientific evidence for relevance rather than reviewing all evidence holistically—which, it is argued, is more consistent with scientific practices.[39] The better option may be to allow the jury to review all methodologically sufficient data and trust the jury to assign appropriate weight to that data based on its relevance as part of a whole body of evidence. This holistic approach allows jurors to consider the cumulative relevance and impact of evidence, and this process may bear greater resemblance to the epistemological endeavors undertaken by scientists or doctors. While theoretically judges could review multiple pieces of evidence simultaneously, it is not clear how the *Daubert* criteria should be applied holistically.

This problem may be exacerbated by courts' tendencies to apply admissibility and sufficiency standards simultaneously, potentially conflating the two concepts.[40] In *Joiner,* the Supreme Court appears to have condoned the consideration of sufficiency concerns in the context of determining admissibility:

> We further hold that, because it was within the District Court's discretion to conclude that the studies upon which the experts relied were not sufficient, whether individually or in combination, to support their conclusions that Joiner's exposure to PCB's contributed

37. *See* Malone & Zwier, *supra* note 35, at 106–07. *Cf.* Christopher B. Mueller, Daubert *Asks the Right Questions: Now Appellate Courts Should Help Find the Right Answers,* 33 SETON HALL L. REV. 987, 989–94 (2003). For two thorough discussions of whether judges or juries are more competent and better suited to review expert evidence, see generally Neil Vidma & Shari Seidman Diamond, *Juries and Expert Evidence,* 66 BROOK. L. REV. 1121 (2001); Joseph Sanders, *The Merits of the Paternalistic Justification for Restrictions on the Admissibility of Expert Evidence,* 33 SETON HALL L. REV. 881 (2003) [hereinafter Sanders, *Paternalistic Justification*]. Sanders's piece is particularly interesting in that it explores naturalized epistemology, "an epistemology that builds on and takes its direction from empirical observations about how we know things." *Id.* at 884. In other words, the validity of a paternalistic justification for excluding evidence from the jury's view should be determined by reference to empirical evidence concerning the jury's ability to convey and understand information. *Id.*

38. *See* Erica Beecher-Monas, *A Ray of Light for Judges Blinded by Science: Triers of Science and Intellectual Due Process,* 33 GA. L. REV. 1047, 1068–69 (1999) [hereinafter *Intellectual Due Process*] (discussing courts' exclusion of individual studies as irrelevant).

39. *Id.* at 1068; David Egilman et al., *Proving Causation: The Use and Abuse of Medical and Scientific Evidence Inside the Courtroom—An Epidemiologist's Critique of the Judicial Interpretation of the* Daubert *Ruling,* 58 FOOD & DRUG L.J. 223, 225 (2003).

40. *See Intellectual Due Process, supra* note 38, at 1072–76 (discussing, inter alia, the Supreme Court's conflation of these two issues in *Joiner*).

to his cancer, the District Court did not abuse its discretion in excluding their testimony.[41]

The dissent in *Joiner* expresses regret that this decision allows courts to consider the sufficiency of individual pieces of evidence rather than consider the weight of the cumulative evidence.[42] Questions of admissibility arising from a challenge under *Daubert* and questions of sufficiency arising from a summary judgment motion are often disposed of in a single hearing.[43] This process arguably facilitates courts' ability to sequentially review the admissibility and sufficiency of each piece of evidence rather than to consider the cumulative impact of all the evidence.[44]

Two other concerns stem from this potential mixing of standards. First, if courts make sufficiency judgments under the guise of admissibility findings, they may skew admissibility standards. For example, evidence may be excluded after a *Daubert* hearing not because such evidence fails to meet the *Daubert* standard, but instead because the judge finds the evidence "insufficient."[45] Second, judges may be able to effectively dismiss cases using sufficiency considerations in the context of a *Daubert* analysis even when they would not be able to do so if forced to address sufficiency considerations directly (for example, when ruling on a motion for summary judgment).[46]

The subtle procedural changes effectuated by the decision may have profound impacts, even outside the law. Post-*Daubert*, defendants may succeed in excluding evidence in a *Daubert* hearing at a higher rate than they would have under *Frye* and subsequently prevail in summary judgment by poking holes in the plaintiff's evidence. Thus, defendants may be less likely to present evidence (specifically studies) of

41. Gen. Elec. Co. v. Joiner, 522 U.S. 136, 146–47 (1997).

42. *Id.* at 152–55 (Stevens, J., dissenting). The dissent also expresses concern that *Joiner* is unfaithful to *Daubert*'s requirement that judges only consider the methodology and not the conclusions of expert's testimony. *Id.* Whether admissibility and sufficiency parallel methodology and conclusions is a debate best saved for another day. *See generally* Carl F. Cranor et al., *Judicial Boundary Drawing and the Need for Context-Sensitive Science in Toxic Torts After* Daubert v. Merrell Dow Pharmaceuticals Inc., 16 VA. ENVTL. L.J. 1, 16–17 (1996) [hereinafter *Judicial Boundary Drawing*]; David L. Faigman, *The Law's Scientific Revolution: Reflections and Ruminations on the Law's Use of Experts in Year Seven of the Revolution*, 57 WASH. & LEE L. REV. 661, 662–65 (2000) (arguing that methodology and conclusions cannot be severed).

43. *See Intellectual Due Process*, *supra* note 38, at 1074.

44. *Id.* at 1075.

45. *See* Daniel Capra, *The* Daubert *Puzzle*, 32 GA. L. REV. 699, 754–55 (1998).

46. *Id.* at 751–55.

their own if such studies are unnecessary and costly.[47] To the extent that defendants are less likely to have to present conflicting evidence, they may be less likely to conduct important studies.[48] Hence, in the toxic torts context, *Daubert* may contribute to the already troubling paucity of knowledge concerning the effects of chemical exposure.

This one-sided presentation of evidence may also exaggerate pre-existing cultural differences between the scientific and legal communities. Judges' analyses of scientific evidence may be affected by their lack of familiarity with the rhetorical styles and cultural values of the scientific community.[49] As discussed below, scientists generally tend not to assert claims with nearly the same degree of certainty as do attorneys. If a judge is reviewing evidence from both parties, she is forced to acknowledge that scientific evidence generally bears this rhetorical style. If instead she is reviewing only the plaintiff's evidence and comparing it with the defendant's legal arguments, she may find the scientific evidence weak and unconvincing and unintentionally be swayed by the discordance in these rhetorical styles.

Moreover, some have argued that one benefit to allowing evidence to go to a jury is that particularly egregious behavior by a defendant may be aired, even if the plaintiff ultimately does not prevail in her case.[50] The plaintiff may gain from the satisfaction of telling her story to a jury, and the public may benefit from an increased awareness of the defendants' actions. Any negative publicity directed toward a defendant may have a desirably deterrent effect on the defendant's harmful behavior.

Finally, federal judges, burdened with heavy dockets, have an incentive to dispose of cases quickly. Often the exclusion of evidence is fatal to the plaintiff. Entrusting judges with this responsibility could result in a greater percentage of exclusions and subsequent grants of summary judgment than might be ideal.

47. Carl F. Cranor & David A. Eastmond, *Scientific Ignorance and Reliable Patterns of Evidence in Toxic Tort Causation: Is There a Need for Liability Reform?*, 64 LAW & CONTEMP. PROBS. 5, 15 (2001) [hereinafter *Scientific Ignorance*].

48. *See id.*

49. *Id.* at 22. *See also* discussion *infra* part II.C.

50. Margaret A. Berger, *Complex Litigation at the Millennium: Upsetting the Balance Between Adverse Interests: The Impact of the Supreme Court's Trilogy on Expert Testimony in Toxic Tort Litigation*, 64 LAW & CONTEMP. PROBS. 289, 324 (2001).

B. Judges Misapply *Daubert* by Placing Excessive Limitations on the Types of Evidence They Will Admit

Critics have charged that some judges have misinterpreted and broadened the reach of *Daubert* by requiring specific types of data and accordingly excluding otherwise valid data that does not conform to these requirements.[51] One prevalent example is in the toxic tort setting, where some judges have viewed *Daubert* to require epidemiological data, and in some instances require that such studies demonstrate a two-fold increase in relative risk.[52] The justification for this requirement is that if normally six out of 1000 people will develop a certain disease and epidemiological data demonstrates that twelve out of 1000 exposed to a particular toxin develop this disease, one can generalize that for any given individual who is both exposed and develops the disease, it is more likely than not true that his or her illness was caused by exposure.[53] At some level, this reasoning is persuasive. Critics, however, point out that other evidence may contribute to establishing the likelihood that the individual's disease was caused by the defendant's product. They further argue that it is inequitable to deny all plaintiffs recovery when some of them (but fewer than half) were in fact injured by the defendant.[54] Additionally, this stringent requirement is based upon the assumption that the background cases of a particular illness and those cases caused by chemical exposure are independent.[55] It is equally likely that a particular injury—cancer, to state the obvious example—is an "endpoint of a pathologic process whose rate is accelerated" by chemical exposure.[56] Exposure may accelerate the onset of a particular injury in many, or even all people.[57] This complicates the relationship between background rates and those found in exposed populations by potentially "underestimat[ing] the degree of risk faced by any individual."[58]

Additionally, such a restriction arguably fails to take into account the eggshell skull principle, which implies that a defendant is still lia-

51. *See generally* TELLUS PUBLICATION, *supra* note 36.

52. *Judicial Boundary Drawing*, *supra* note 42, at 33–40.

53. *Id.* at 37–38.

54. *See* Mark Geistfeld, *Scientific Uncertainty and Causation in Tort Law*, 54 VAND. L. REV. 1011, 1035 (2001).

55. *Id.* at 1033–34.

56. *Id.*

57. *Id.*

58. *Id.*

ble to an abnormally sensitive or predisposed plaintiff.[59] To remain consistent with this tort principle, some argue that *Daubert* should allow for the admission of methodologically sound evidence. This would demonstrate that although exposure to agent X is not likely to cause harm to the average person, the plaintiff was particularly susceptible to this particular harm, and in the expert's opinion the exposure was more likely than not the cause of the plaintiff's illness.[60]

More significantly, scientists and lawyers question the basic assumption that epidemiological data is necessary to show causation.[61] They emphasize that scientists do not specifically require this type of data to establish causation.[62] Such a requirement disadvantages plaintiffs given the dearth of epidemiological data, difficulty and cost of conducting such studies, and shortcomings of existing studies.[63] This argument also applies to other types of evidence that have in the past been excluded by certain courts, including animal studies[64] and differential diagnosis.[65]

As addressed earlier, judges often apply the *Daubert* factors to individual pieces of evidence instead of considering the relevance and reliability of the body of proffered evidence as a whole. Once again, reviewing the cumulative weight of evidence may provide for more accurate assessments of the relevance and reliability of evidence, at least insofar as this is the means by which scientists often assess such evidence. While some would argue that a piecemeal analysis is supported by *Daubert* and endorsed by *Joiner*, a holistic analysis of evidence may not be inconsistent with the decision.

C. *Daubert* Offers a Flawed Concept of "Good" Evidence or Incomplete Instructions on How to Identify It

Much criticism has been directed toward the content of the decision itself. As referenced above, critics argue that *Daubert* encourages a piecemeal analysis of evidence and a checklist approach to individual

59. *Scientific Ignorance, supra* note 47, at 44–45. The eggshell skull rule is "the principle that a defendant is liable for a plaintiff's unforeseeable and uncommon reactions to the defendant's negligent or intentional act." BLACK'S LAW DICTIONARY 218 (Pocket ed. 1996).

60. *Id.*

61. *Judicial Boundary Drawing, supra* note 42, at 32.

62. *Id.*

63. *See generally Scientific Ignorance, supra* note 47.

64. *Id.* at 27.

65. *See generally* Joseph Sanders & Julie Machal-Fulks, *The Admissibility of Differential Diagnosis Testimony to Prove Causation in Toxic Tort Cases: The Interplay of Adjective and Substantive Law*, 64 LAW & CONTEMP. PROBS. 107 (2001).

studies.[66] They express concern that the admissibility criteria embraced by the decision are oversimplified, but judges nonetheless may be tempted to consider the factors to be a cookbook recipe for good science.[67] Scholars also complain that the decision encourages a sequential instead of cumulative assessment of evidence.[68] They argue that a particular study may not, for example, be statistically significant at the .05 level due to a small sample size, but when considered in tandem with additional studies or different types of evidence, may contribute to a persuasive body of evidence.[69] According to one publication, good science proceeds by "assessing totality of the evidence," not by conducting a separate analysis of individual factors.[70] Scholars are concerned that judges will apply the reliability criteria in a formulaic manner. Those criteria, once again, are: (1) whether the evidence is "falsifiable" and has been tested, (2) peer review and publication, (3) the technique's known or potential rate of error, (4) whether standards for controlling the technique's operation exist, and (5) the general acceptance of the methods and reasoning.[71]

Critics also charge that *Daubert*, like many multi-factored tests, is too flexible and leaves too much discretion in the hands of the judge.[72] Its application is likely to produce inconsistent, arbitrary, and unpredictable results.[73] This potential problem is exacerbated by the fact that the decision only gently recommends use of the five reliability factors (as opposed, in some sense, to the converse concern that judges will use these factors in a cookbook type fashion). Furthermore, the actual factors chosen by the court may be faulted. *Daubert* has been criticized for espousing an over-reliance on peer review and statistical significance.[74] Recent scholarship has suggested that the

66. *See Judicial Boundary Drawing, supra* note 42, at 25–26.

67. *Id.* at 21–27.

68. TELLUS PUBLICATION, *supra* note 36, at 7.

69. *See Judicial Boundary Drawing, supra* note 42, at 34. Statistical significance is the "probability that the observed relationship (e.g., between variables) or difference (e.g., between means) in a sample occurred by pure chance ("luck of the draw") and that in the population from which the sample was drawn, no such relationship or difference exists." StatSoft, Elementary Concepts in Statistics, http://www.statsoft.com/textbook/esc.html (last visited Oct. 27, 2005). Researchers often require at a minimum that a study be significant at the .05 level. This means that there is only a 5% probability that the relationship between variables or means is just the result of chance. *Id.*

70. *Id.*

71. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 593–95 (1993).

72. *See Judicial Boundary Drawing, supra* note 42, at 5.

73. *See id.*

74. Joelle Anne Moreno, *Eyes Wide Shut: Hidden Problems and Future Consequences of the Fact-Based Validity Standard*, 34 SETON HALL L. REV. 89, 95 (2003).

peer review process may be faulted and may not, in fact, be an accurate indication of "good science."[75] Statistical significance, particularly at the .05 level, is ideal but may not be a necessary or practical standard.[76]

Daubert is also subject to the criticism that it fails to adequately address the problem of causal uncertainty in the toxic tort context.[77] Given that our knowledge about the chemical universe is severely limited and the mechanisms by which chemicals cause illness are largely unknown, it may not be appropriate to treat causation in toxic tort cases in the same manner as it is treated in other tort contexts.[78] The Court could have addressed this problem by adopting a probabilistic notion of causation—for example—by admitting evidence demonstrating a less than 50% chance of causation and permitting juries to give partial awards based on the probabilistic likelihood of causation. Alternatively, the Court could have required that once a plaintiff makes a showing of a scintilla of evidence, the burden of proof shifts to the defendant.[79] This could compensate for the burden otherwise placed on plaintiffs who lack the resources to conduct their own studies and thus overcome the current state of ignorance concerning the effects of most chemicals. As discussed below, this issue of causation exposes larger conflicts between the cultures and aims of science and the law.

Finally, *Daubert* has been criticized for incorporating a faulty philosophy of science. Some critics debate whether the decision incorporated a view of science as an objective enterprise or as a social and cultural phenomenon.[80] Others charge that the decision struck an incoherent synthesis between these differing views on the true nature of science.[81] On a more positive note, some scholars argue that *Daubert* is sufficiently flexible to allow judges to recognize the social, institutional, and rhetorical aspects of science and not focus on an idealized

75. *Id.*

76. *See* Egilman, *supra* note 39, at 236–39.

77. *See Scientific Ignorance*, *supra* note 47, at 46.

78. *See generally id.* See *David Rosenberg, The Causal Connection in Mass Exposure Cases: A "Public Law" Vision of the Tort System*, 97 HARV. L. REV. 851 (1984), for an excellent discussion of this problem (published prior to *Daubert*).

79. *Scientific Ignorance*, *supra* note 47, at 47.

80. David S. Caudill & Richard E. Redding, *Junk Philosophy of Science?: The Paradox of Expertise and Interdisciplinarity in Federal Courts*, 57 WASH. & LEE L. REV. 685, 691–92 (2000).

81. Margaret G. Farrell, Daubert v. Merrell Dow Pharmaceuticals, Inc.: *Epistemology and Legal Process*, 15 CARDOZO L. REV. 2183, 2198–207 (1994).

version of it when rendering their decisions.[82] In practice, judges sometimes may fall prey to an idealized notion of science (perhaps driven by a strict adherence to the five *Daubert* factors), but select cases provide examples of judges taking a more "pragmatic" approach towards science, one that arguably provides a better mirror of how scientists actually operate.[83] Unfortunately, a thorough discussion of this issue could dwarf the remainder of this Article and thus should be left for another day.

D. The Best Description of Good Science Is Not the Best Description of Good Scientific Evidence

Emerging from recent scholarship is a thread of criticisms that addresses the conflicting cultures of science and the law and asks whether applying the scientific community's best description of good science to evidence presented in a court of law necessarily and undesirably entails importing scientific values into the judicial sphere.[84] This general sentiment is that "[i]f scientific approaches to distributions of mistakes or conceptions of accuracy dominate in the law, they risk distorting the law's goals."[85] These criticisms may, of course, be applicable to any standards for the admissibility of evidence. In deciding *Daubert* as it did, the Supreme Court may have missed an opportunity to develop a formula for translating scientific knowledge into the context of litigation without carrying over certain characteristics of science that create biases in the courtroom.

Carl Cranor, a professor of philosophy at University of California, Riverside, has observed that science (particularly toxicology) and the law place different values on two types of errors: false positives and false negatives.[86] False positives are erroneous findings of causation when none, in fact, exists. False negatives are erroneous findings of no causation when one does exist. Tort norms aim to achieve a balance between errors that benefit the plaintiff and those that benefit the

82. *See generally* David S. Caudill & Lewis H. LaRue, *Why Judges Applying the* Daubert *Trilogy Need to Know About the Social, Institutional, and Rhetorical—And Not Just the Methodological—Aspects of Science*, 45 B.C. L. REV. 1 (2003).

83. *Id.*

84. *See generally Scientific Ignorance*, *supra* note 47.

85. *Id.* at 19. *Cf.* Bert Black, Francisco J. Ayala & Carol Saffran-Brinks, *Science and the Law in the Wake of* Daubert*: A New Search for Scientific Knowledge*, 72 TEX. L. REV. 715, 753 (1994) ("We believe lawyers and judges will accomplish the task of understanding science best if they . . . rely on the same process of review that scientists use.").

86. *Scientific Ignorance*, *supra* note 47, at 19–25.

defendant.[87] In other words, tort law is indifferent as to whether the plaintiff erroneously recovers (false positive) or the defendant erroneously is found not liable (false negative).[88] The hope is that a relative number of each will approximately cancel each other out. In contrast, scientists and scientific bodies, through culturally regulated behavior (peer review, etc.), and because of the degree of statistical significance generally demanded of studies, place a greater emphasis on avoiding false positives.[89] This effect is due largely to the manner in which scientists set error rates in experiments (generally allowing a greater number of false negatives in order to meet strict limits on the percentage of false positives).[90] The result is that when scientific studies are introduced into the legal context, they bring along an inherent bias against finding a causal link between a chemical agent and physical harm.[91]

Another source of tension identified by Cranor concerns rhetorical differences between science and the law.[92] While law embraces the ideal of the zealous advocate, the rhetorical style adopted by scientists could be characterized as hedging.[93] While good litigators express a strong and genuine-sounding commitment to their arguments, scientists often exhibit a limited commitment to claims. The result is that when judges are called upon to consider the adequacy of a party's scientific claims, they may make an inaccurate assessment because the evidence may sound comparatively unconvincing. In theory this problem should affect evidence presented by both the plaintiff and the defendant equally and hence should balance out.[94] However, as discussed above, it is plausible that under *Daubert*, defendants are presenting considerably less evidence than they previously did, instead relying on attacks to the plaintiff's evidence. If this is the case, then judges' culturally-based disinclination to find scientific evidence persuasive may regularly work to the disadvantage of plaintiffs.

87. *Id.* at 19–20.

88. *Id.* at 19.

89. *Id.* at 20.

90. *Id.*

91. *Id.*

92. *Id.* at 18–24.

93. *Id.* at 22–23.

94. For a discussion of the reverse problem, where judges are too persuaded by the authoritative force of science, see Caudill & LaRue, *supra* note 82, at 43–48.

Critics are also concerned by inconsistencies in how lawyers and scientists understand causation.[95] The law demands certainty and finality and generally forces its participants to render decisions and verdicts in binary pairs, such as causation/no causation and liability/no liability, even where evidence is ambiguous, uncertain, complex, and immature.[96] In contrast, the scientific enterprise generally embraces probability and uncertainty and does not require the occasionally premature dichotomization of outcomes often required by the courts. As Cranor points out, medical scientists find the question of individual causation unanswerable, and scientists generally are reticent to reach absolute conclusions.[97] Statutes of limitations, the principle of res judicata, and the need to promptly compensate sick plaintiffs limit the court's ability to accommodate the slow accumulation of sometimes inconsistent scientific data. Furthermore, one scholar has proposed that while *Daubert* captured an accurate view of how scientists approach the problem of "deciding which information to consider when deciding questions of scientific fact," this view necessarily and frequently leads to uncertainty with regard to causation.[98] *Daubert* may have been a missed opportunity at addressing these problems, (though, admittedly, it is not clear that an evidentiary standard is the best means by which to address these problems).

E. Judge-Made Rules Under *Daubert* Violate the Principles of *Erie*

Brooklyn Law School Professor Margaret Berger has argued that judge-made laws under *Daubert*, such as the exclusion of animal studies or epidemiological data that demonstrates a relative risk of less than 2.0, may violate the principles of *Erie Railroad v. Tompkins*[99] when applied in a federal court sitting in diversity jurisdiction.[100] In *Erie*, the Supreme Court decided that a federal court sitting in diversity jurisdiction may apply federal procedural law, but must apply state substantive law.[101] The Court's subsequent decision in *Hanna v. Plumer*[102] held that where a federal rule is adopted under the Rules Enabling

95. Joëlle Anne Moreno, *Beyond the Polemic Against Junk Science: Navigating the Oceans that Divide Science and Law with Justice Breyer at the Helm*, 81 B.U. L. REV. 1033, 1060–65 (2001).

96. *See id.* at 1062.

97. *Scientific Ignorance*, *supra* note 47, at 7–8.

98. Hedi Li Feldman, *Science and Uncertainty in Mass Exposure Litigation*, 74 TEX. L. REV. 1, 2 (1995).

99. 304 U.S. 64 (1938).

100. Berger, *supra* note 50, at 291.

101. *Erie*, 304 U.S. at 78.

102. 380 U.S. 460 (1965).

Act[103] or a statute authorized by Congress, federal judges are "to follow federal practice even in the face of a contrary state rule, unless the federal practice 'transgresses . . . the terms of the Enabling Act [or a] constitutional restriction.'"[104] As to a rule adopted pursuant to the Rules Enabling Act, *Erie* would only be violated if such a rule violated the Act's prohibition that a rule "not abridge, enlarge or modify the substantive rights of the litigant."[105]

However, Berger points out that when judges establish black-letter policy rules, excluding animal studies, for example, they are no longer applying the text of the rule as in *Hanna*, but instead are moving "considerably beyond the scope of Rule 702."[106] Federal Rule 702 nowhere requires the exclusion of animal studies or, for that matter, a doubling of risk.[107] In a subsequent case, *Gasperini v. Center for Humanities, Inc.*,[108] the Supreme Court clarified that the virtually automatic acceptance of federal rules dictated in *Hanna* does not apply to federal judge-made procedural rules.[109] The Court went on to say that if a state practice exists that conflicts with a federal practice (not with the text of a federal rule), the lower court must consider whether the state version is substantive or merely procedural.[110] If it is substantive, the lower court must apply the state version.[111] Berger then explains that while it is difficult to determine which of a series of tests the court should apply to decide whether a state practice is substantive, each of the tests "points in the direction of applying state law when the issue concerns a policy-based rule on proving causation."[112]

In light of this jurisprudence, Berger is persuaded that "[t]he principles that the Court has at various times identified as underlying *Erie* and its progeny all point to a need to consider state law to determine whether a federal court may apply its own judge-made rule on the admissibility of expert proof about causation in a toxic tort action."[113]

103. 28 U.S.C. §§ 2071–77 (2000).
104. Berger, *supra* note 50, at 310.
105. *Id.* at 311.
106. *Id.* at 312.
107. *Id.*
108. 518 U.S. 415 (1996).
109. *See* Berger, *supra* note 50, at 312–13.
110. *Id.*
111. *Id.* at 313.
112. *Id.*
113. *Id.* at 316.

III. Empirical Data Regarding *Daubert*'s Effects

This Part reviews the findings of recent studies designed to determine how *Daubert* has affected admissibility determinations. These studies not only help answer questions about how judges are applying the decision and how parties to litigations have responded to this new standard, but also shed light on criticisms of *Daubert*.

A. RAND Study Reflects Judicial Scrutiny and the Limited Role of Reliability Factors

In 2001, the RAND Institute for Civil Justice issued a report analyzing trends in 399 federal district court opinions issued between January 1980 and June 1999.[114] This study has been widely cited as providing some of the best evidence that *Daubert* indeed heightened admissibility standards. By reviewing pre- and post-*Daubert* decisions, the researchers were able to investigate whether, and in what ways, admissibility decisions have been influenced by *Daubert*. Overall, the study showed numerous changes in the wake of *Daubert* that the authors believe are not merely the impact of general trends in litigation or changes in the composition of the federal bench.[115]

Following *Daubert* the authors found a significant rise in the proportion of evidence excluded.[116] They noted an increase in the proportion of rulings on challenges discussing reliability, suggesting that parties were more frequently challenging evidence based on its reliability.[117] The success rate of these challenges increased as well.[118] It appears that judges were scrutinizing reliability more carefully, applying stricter standards and, as a result, excluding an increasing proportion of evidence.[119] For example, the study found that for evidence based on physical science in products liability cases in the Third Cir-

114. RAND Study, *supra* note 3, at xiii. The RAND Institute for Civil Justice is an independent research program within the RAND Corporation, a non-profit think tank.

115. *Id.* For example, the authors found that following the decision, judges initially started scrutinizing physical science-type evidence more closely, the signature trend that would be expected if they were basing their review on the *Daubert* decision. The authors came to the conclusion that *Daubert* was at least partially responsible for subsequent changes in admissibility despite finding an increase in challenges to expert testimony just before the decision was rendered, possibly suggesting the beginning of a trend that predated *Daubert* (however, this finding was not statistically significant). RAND Study, *supra* note 3, at xv.

116. RAND Study, *supra* note 3, at 55.

117. *Id.* at xv.

118. *Id.*

119. *Id.* at xv–xvi.

cuit, the exclusion rates spiked to 70% between mid-1995 and mid-1996, up from 53% for the two years preceding *Daubert*.[120]

To determine whether this apparently greater scrutiny of reliability is the result of substituting terms (assessing "reliability" instead of an expert's "qualifications," for example) rather than an actual change in how carefully judges are reviewing evidence overall, the authors looked for trends in how judges evaluate relevance and qualifications as well as reliability.[121] If judges were simply switching terminology, one would expect to see a decrease in decisions discussing relevance or qualifications to account for the increase in discussion of reliability. Instead, the researchers found trends in successful challenges based on relevance and qualifications that paralleled that of reliability,[122] leading them to conclude that judges are scrutinizing testimony more carefully with regard to all criteria.[123]

Interestingly, this trend changed in 1997, when the percentage of challenges and their success rate started to decline.[124] Post-1997, the study shows a gradual decrease in the percentage of challenged evidence found unreliable.[125] The authors interpret this data to reflect changing behavior by parties in response to the tightening of standards, specifically inferring that "parties proposing evidence either did not propose or withdrew evidence not meeting the new standards, or better tailored evidence they did propose to fit the new standards."[126] They also hypothesize that parties challenging evidence may have been so encouraged by past success that they cast a wider net in terms of what evidence they challenged, and hence the proportion of successful challenges may have dropped.[127]

Daubert appears to have had a significant impact on motions for summary judgment: summary judgment was requested and granted at a considerably higher frequency following the decision. It was requested in one-third of challenges to evidence and granted half the time.[128] This trend was particularly fatal for cases, with 88% of motions made in challenges to evidence proposed by plaintiffs and nearly

120. *Id.* at xvi.
121. *Id.* at xv.
122. *Id.* at 49–50.
123. *Id.* at 52.
124. *Id.* at xvii–xviii.
125. *Id.* at xvii.
126. *Id.*
127. *Id.*
128. *Id.* at 56.

90% of judgments coming down against plaintiffs.[129] The success rate of such challenges does appear to have fallen off slightly in the latter years of the study.[130]

Beyond the oft-cited general observations above, the RAND Study also contains some interesting specific findings that bear on the mechanisms of how judges apply *Daubert*. The first finding concerns the role of general acceptance in judicial evaluations for reliability. Despite criticisms that *Daubert* tightened admissibility standards, the decision itself appeared to liberalize the standard by demoting general acceptance to one of five factors to be considered.[131] Thus, it appeared to allow methodologically sound minority opinions. Nevertheless, the actual effect of the decision has been largely to the contrary. The RAND Study found that before *Daubert*, general acceptance was not commonly used as a factor for admissibility, but when it was addressed, general acceptance was usually sufficient for admissibility.[132] After the decision, general acceptance was no longer sufficient.[133] These findings are not necessarily inconsistent with *Daubert* in that the decision reduces the significance of general acceptance. The surprising finding, however, was that after *Daubert*, a lack of general acceptance was as much a barrier to admission as before, and possibly a greater one.[134] Using regression analysis, the authors found that general acceptance pre-*Daubert* was not a good indicator of whether evidence would be found reliable or not, but general acceptance post-*Daubert* was, in fact, a good indicator of reliability.[135] According to the authors of the study, there is "[n]o indication it became easier for novel evidence to be admitted."[136] This seems contrary to *Daubert*'s thrust toward a greater emphasis on methodology and procedure and decreased emphasis on the majority viewpoint.

Another revealing finding in the RAND Study is the frequency with which specific factors were discussed in decisions on admission. Following *Daubert*, all five factors were addressed more frequently.[137] The trend in the discussion of these factors roughly parallels the trend in the discussion of reliability generally: both increased significantly

129. *Id.* at 56, 62.
130. *Id.* at 56.
131. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 593–95 (1993).
132. RAND Study, *supra* note 3, at 44.
133. *Id.*
134. *Id.*
135. *Id.* at 45.
136. *Id.*
137. *Id.* at 40.

after the decision, peaked, and decreased starting around 1997.[138] Interestingly, there was also an increase in discussion of non-*Daubert* factors after the decision, and this discussion continued to increase in frequency throughout the entire time period analyzed.[139] Thus, while the frequency of *Daubert* factors eventually declined, the frequency of non-*Daubert* factors did not. One factor in particular, "clarity and coherence of underlying method and procedures," showed a continual and dramatic increase in frequency of discussion.[140] The authors' interpretation of these findings is that immediately after the decision, judges, acting prudently, stuck to the listed factors, but once they gained greater familiarity with the role of scrutinizing evidence, they became more confident in their freedom and ability to consider other factors.[141]

Finally, as discussed above, in the wake of *Daubert*, judges applied greater scrutiny to factors other than reliability, even though the decision did not change the standards for such review. Judges more frequently discussed and excluded evidence on the basis of relevance and qualifications,[142] increasingly examined all evidence (not just hard science),[143] and increasingly focused on theory, methods, and procedures underlying expert evidence.[144]

B. Federal Judicial Center Study Results Affirm RAND Study Findings

A second comprehensive study on the impacts of *Daubert* was conducted by the Federal Judicial Center.[145] The researchers conducted three surveys: two were distributed to federal judges in 1991 and 1998, and one was distributed to attorneys in 1999.[146] The three surveys contained questions concerning the judges' and attorneys' general exper-

138. *Id.* at 42.

139. *Id.* at 43.

140. *Id.*

141. *Id.* at 40–41.

142. In the authors' words, "[i]t appears that once judges started acting as more active gate keepers, they more carefully examined relevance, qualifications, and other considerations for admitting evidence, in addition to reliability." *Id.* at 61–62.

143. *Id.* at 63.

144. *Id.* at 62.

145. FJC Study, *supra* note 3. The Federal Judicial Center is the education and research agency for the federal courts.

146. *Id.* at 311.

iences with expert testimony and their specific experiences in recent civil trials.[147]

Notably, in reviewing the selected cases, the authors found that the percentage of trials containing experts testifying for plaintiffs and the percentage containing experts testifying for defendants remained relatively consistent from 1991 to 1998.[148] The mean number of experts testifying decreased slightly, but limits to the 1991 data prevented the authors from deciphering whether this reduction in experts was particular to defendants or plaintiffs.[149]

With regard to admissibility generally, the results of this study are consistent with those of the RAND Study. In 1998, judges reported that they were more likely to scrutinize expert testimony and that they were less likely to find it admissible.[150] They reported limiting or excluding expert testimony in 41% of the 1998 cases, versus 25% of the cases referenced in the 1991 survey.[151] As the authors point out, the surveys only contained questions about trials that contained some expert testimony and did not capture information on cases in which all expert testimony was ruled inadmissible.[152] Thus, the true rate of expert testimony exclusion may be higher (particularly given the increase in *in limine* motions, discussed below).[153] Furthermore, 33% of judges in 1998 reported that they admitted evidence less frequently than they did prior to *Daubert*.[154] Sixty-one percent of attorneys agreed that judges were admitting less evidence.[155] This assessment is supported by the finding that "[j]udges who ruled on expert admissibility issues in the cases sampled for the recent survey permitted 59% of cases to proceed to trial without limitation on the evidence, whereas judges who ruled on admissibility in the early survey permitted a full 75% of cases to proceed without limitation."[156]

In *Daubert*'s wake, motions *in limine* (resulting in pre-trial or "*Daubert*" hearings) have been used more frequently as a device for addressing admissibility.[157] According to the 1991 study, prior to the

147. *Id.* at 309. Judges were asked about their most recent civil trial and the attorneys involved were questioned about it as well. *Id.*
148. *Id.* at 318–19.
149. *Id.* at 319.
150. *Id.* at 322.
151. *Id.*
152. *Id.*
153. *Id.*
154. *Id.* at 329.
155. *Id.*
156. *Id.* at 330.
157. *Id.* at 321.

decision, admissibility questions were most often raised at trial.[158] By contrast, according to the 1998 survey, admissibility issues arose most frequently in the context of motions *in limine*.[159] Judges reported regular use of *Daubert* hearings on admissibility of evidence, with 77% reporting that they "commonly held a *Daubert* hearing on evidence admissibility,"[160] up from 51% in the 1991 survey.[161] Furthermore, 45% of judges reported that they were hearing pre-trial motions more frequently than they had been pre-*Daubert*, and 56% of attorneys agreed.[162] In light of these findings, the authors were surprised to discover that 60% of judges stated that their use of procedures had not changed very much since *Daubert*.[163] The authors conclude, nonetheless, that "*Daubert* gives every appearance of having affected the judicial approach to handling expert evidence in federal civil cases."[164]

More than half of attorneys surveyed reported making greater uses of motions *in limine* to exclude testimony.[165] A significant number of attorneys reported making more objections to admissibility of expert testimony at trial (41%) and also making more motions for summary judgment (24%).[166] This is supported by the finding that after *Daubert* there was a significant increase in the frequency with which judges ruled on admissibility following a motion *in limine*, from 32% to 72%.[167]

Overall, the study strongly supports the hypothesis that since *Daubert* parties are challenging expert evidence more frequently and at an earlier stage in litigation. Judges are excluding more evidence, and parties are making more motions for summary judgment. As with the RAND Study, it is useful to look at some of the findings in detail to better understand the mechanism by which *Daubert* has caused these changes.

158. *Id.*

159. *Id.* When admissibility was raised, in 72% of the cases it arose in a motion *in limine*. However, such issues still often emerged in response to an objection made at trial—64% of cases contained admissibility questions raised at the trial phase. *Id.*

160. *Id.* at 326 (although 29% of judges indicated that they only held such hearings in cases with "difficult scientific evidence"). *Id.*

161. *Id.* at 327.

162. *Id.* at 329.

163. *Id.* Possible explanations for this discrepancy are that judges view these changes simply as a matter of degree, that the increase in hearings was caused by the 40% who reported changes in management of expert testimony, or that judges are poorly assessing changes in their procedures. *See id.*

164. *Id.*

165. *Id.*

166. *Id.*

167. *Id.*

Interestingly, the authors note that while 18% of the judges who excluded evidence based their decision in part on a finding that the methods and principles of the expert were unreliable, judges rarely discussed the specific *Daubert* criteria for reliability.[168] In cases where evidence was excluded, judges identified problems with general acceptance, peer review, and insufficient theory testing less than 8% of the time.[169] Issues with falsifiability and error rates were discussed in less than 2% of these cases.[170]

The study also found that the reasons for limiting or excluding testimony have not changed significantly in the wake of *Daubert*.[171] In 1998, as in 1991, most exclusions were based on the judge's assessment that the evidence was "not relevant, the witness was not qualified, or the testimony would not have assisted the trier of fact."[172]

C. Groscup Study

In 2002, Jennifer Groscup—now a professor at John Jay College of Criminal Justice—along with other professors and researchers, published a study looking at the effects of *Daubert* on federal and state criminal cases.[173] The researchers reviewed appellate decisions, coding for over 100 variables related to the admissibility of evidence.[174] They recorded the length of discussion of some of these variables and coded for the influence of select factors on the decisions.[175] In addition, they noted whether the testimony did or did not meet certain criteria.[176]

As indicated, this study only concerned *criminal* cases,[177] whereas this Article focuses on civil cases.[178] While the findings discussed be-

168. *Id.* at 323.
169. *Id.*
170. *Id.*
171. *Id.* at 330.
172. *Id.*
173. Groscup Study, *supra* note 3.
174. *Id.* at 342–44. As indicated, *supra* note 22, some scholars do not recognize a fifth factor.
175. *Id.* at 343.
176. *Id.*
177. *Id.* at 342.
178. In contrast to the RAND and FJC Studies, the Groscup Study found no change in the basic rates of admission at the trial and appellate court following *Daubert*. *Id.* at 345. This finding is somewhat surprising in light of the studies above, but likely is the result of differences between the criminal and civil settings and may be complicated by the fact that investigators only reviewed cases on appeal. *See id.* at 345–46. The authors note that:

> One explanation for the lack of any changes in the observed rates of admission before versus after *Daubert* is that admissibility depends on the party offering the

low are therefore of limited value in determining how *Daubert* impacted admissibility in civil cases, some of the key results of this study are consistent with those found in the two studies of civil cases discussed above. The Groscup Study, therefore, lends some support to the patterns of judicial application of *Daubert* identified above.

Consistent with the studies of civil cases, the authors of this study noted that the "most mysterious impact of the *Daubert* decision is the lack of discussion devoted to the four *Daubert* criteria themselves."[179] They found no increase in the length of discussion devoted to falsifiability, peer review, or error rate, and a decrease in the length of discussion of general acceptance.[180] The authors observed that these factors are "rarely given more than a cursory mention" in appellate decisions[181] and that judges "rarely made statements that the four *Daubert* criteria were or were not met."[182] The most interesting finding in this study was the outcome of a model the authors developed to determine which judgment criteria best predict whether evidence will be deemed admissible. The authors failed to find a significant relationship between the four *Daubert* factors and admissibility.[183] Instead, the most predictive factors were whether the testimony assists the trier of fact, is not regarded as prejudicial, is regarded as relevant, is regarded as reliable, and whether the expert is deemed qualified.[184] This finding suggests that the courts appear to be relying on the re-

testimony. The party for whom the key expert testified was significantly related to admission at both the trial court, $X^2(1, N = 686) = 512.70$, $p < .001$, and the appellate court levels, $X^2(1, N = 670) = 251.79$, $p < .001$. At both adjudicative levels, experts proffered by the prosecution were more likely to be admitted than experts proffered by defendants.

Id. at 346. See also D. Michael Risinger, *Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left on the Dock?*, 64 ALB. L. REV. 99 (2000), for a discussion of the differences in outcomes of challenges by civil versus criminal defendants.

179. Groscup Study, *supra* note 3, at 353.

180. *Id.* As an interesting side note, the authors did observe that decisions contained lengthier discussions of error rate and general acceptance for cases involving scientific experts than for those involving other types of experts. *Id.*

181. *Id.* at 367. In contrast, discussion of *Daubert* generally was "lengthy." *Id.* at 365. The authors also note that there was a significant drop off in discussion and importance rating of general admissibility. *Id.* at 367. This is surprising given its increased influence noted in the RAND Study. They speculate that after *Daubert*, judges may just be less willing to discuss general acceptance, even if it does factor into their decisions. *Id.* at 366–67.

182. *Id.* at 359.

183. *Id.* at 353–56.

184. *Id.* at 363. While the authors did observe an increase in influence for falsifiability, peer review, and error rate, they also noted that "the criteria with the highest ratings of influence were derived from the Federal Rules of Evidence, namely assisting the trier of fact, expert qualifications, relevance of the testimony, and potential for prejudicial impact . . . ." *Id.* at 354.

quirements of the Federal Rules of Evidence more than the *Daubert* factors.[185]

Overall, the authors conclude that while *Daubert* has caused judges to take on the responsibility of gatekeepers, judges have responded by applying the requirements of the Federal Rules of Evidence more stringently rather than utilizing the factors identified in the *Daubert* decision.[186]

D. Cheng & Yoon Study Considers Whether a State's Choice of *Frye* or *Daubert* Has Any Practical Significance

Edward Cheng and Albert Yoon took a different approach.[187] They focused on state courts, some of which adopted the *Daubert* standard and others of which adhere to the old *Frye* standard. Cheng and Yoon asked whether a state's adoption of one standard or another actually matters. To answer this question, they took the novel approach of looking at removal rates from state to federal court in tort cases in order to measure litigants' perceptions of whether they will fare better under one standard or the other.

Cheng and Yoon first looked at a narrow region, considering removal rates in Connecticut and the Eastern District of New York ("EDNY"). In 1997, Connecticut adopted the *Daubert* standard, whereas New York continues to adhere to *Frye*.[188] Using data from the EDNY as a control group, the authors analyzed whether removal rates in Connecticut changed when the state moved from *Frye* to *Daubert*.[189] Somewhat surprisingly, they found no statistically significant changes correlating with this adoption of *Daubert* over *Frye*.[190]

Expanding this study to sixteen states, they found similar results.[191] Accounting for year to year and state to state variations, Cheng and Yoon determined statistically that "the DAUBERT variable—whether a state follows the Daubert standard in the year in question—has a vanishingly small effect on removal rate. DAUBERT contributes only five-thousandths of a percentage point to a state's re-

185. *Id.* at 354. The authors note, however, that the courts paid more attention to the *Daubert* factors when assessing scientific evidence, and caution that parties presenting scientific evidence should still address these factors. *Id.* at 369.

186. *Id.* at 370.

187. Cheng & Yoon Study, *supra* note 4.

188. *Id.* at 475.

189. *Id.* at 486.

190. *Id.* at 488.

191. *Id.* at 498.

moval rate, and the result is not statistically significant."[192] This effect remained even when the authors tweaked various factors to account for the differing sizes of jurisdictions' caseloads.[193]

Finding that a state's choice of evidentiary standards does not have a statistically significant effect on removal rates in that state,[194] the authors inferred that a "state's adoption of *Frye* or *Daubert* makes no difference in practice."[195] By the authors' own admission, this assertion relies on at least three bulky assumptions: (1) that procedural changes impact defendants' decisions to remove (in some way that would actually be captured by this analysis), (2) that defense counsel can accurately predict the "practical ramifications of the scientific admissibility standard adopted in his or her jurisdiction,"[196] and (3) that the results are not skewed by cases in which the case fails to meet the amount in controversy requirement ($75,000) for removal.[197]

Though they acknowledge the significance of these assumptions, Cheng and Yoon conclude that their study lends support to the theory that the actual doctrinal test enumerated in *Daubert* is substantially less important than the awareness the decision generated.[198] They assert:

> The results of this study are consistent with the theory that the power of the Supreme Court's *Daubert* decision was not so much in its formal doctrinal test, but rather in its ability to create greater awareness about the problems of junk science. . . . [C]ourts apply some generalized level of scrutiny when considering the reliability of scientific evidence, regardless of the governing standard. If accepted, this thesis suggests that debates about the practical merits and drawbacks of adopting a *Frye* versus a *Daubert* standard are largely superfluous.[199]

Cheng and Yoon then recommend that attempts at changing or improving the way courts treat scientific evidence could be more effective if "advocates for rigorous use of scientific evidence shifted their focus away from tinkering with doctrinal tests and instead toward 'softer' solutions that increase the judiciary's understanding of scientific concepts and processes."[200] One such example of a softer solution

192. *Id.*

193. *Id.* at 499.

194. *Id.* at 503.

195. *Id.*

196. *Id.* at 508.

197. *Id.* at 506–08.

198. *Id.* at 503–09.

199. *Id.* at 503. The authors point out in a footnote that theoretical discussions about the differences still have value. *Id.* at 503 n.62.

200. *Id.* at 504.

is focusing more on how judges are educated.[201] Ultimately, the authors conclude that taken against the background of numerous studies showing that *Daubert* did have a substantial impact on how scientific evidence is treated, their study "strongly suggests that *Daubert*'s influence was not from its doctrinal reform, but from its educative function."[202]

E. Other Studies

Numerous other studies have focused on aspects of judicial gatekeeping in the wake of *Daubert*. Much literature has been written addressing the proficiency of judges and jurors in assessing the quality of expert evidence. Probably the most cited study is Sophia Gatowski's survey of 400 state court trial judges, which found that 82% of judges demonstrated a clear understanding of general acceptance and 71% of judges understood the utility of the peer review process. As to the peer review process, only 4% of judges demonstrated a clear understanding of falsifiability and of error rates.[203] Furthermore, 73% of judges had no experience with epidemiological data. In addition, 96% of judges reported that they had not received instruction about general scientific methods and principles (even though the majority had received some CLE training).[204] Unfortunately, a thorough discussion about the relative abilities of judges and jurors to separate bad from good evidence is beyond the scope of this Article, but it bears noting that some studies have cast considerable doubt on the frequent assumption that judges necessarily are superior in this task.[205]

Another very interesting study is Michael Risinger's comparison of the outcomes of challenges to expert testimony in civil and criminal cases.[206] Most significantly, Risinger found that in federal appellate and district court opinions as well as state court decisions, civil defendants prevail in their challenges to expert testimony most of the time, while criminal defendants "virtually always lose their reliability chal-

201. *Id.*

202. *Id.* at 505.

203. Neil Vidmar & Shari Seidman Diamond, *Juries and Expert Evidence*, 66 BROOK. L. REV. 1121, 1172–73 (2001) (discussing Sophia Gatowski et al., *Asking the Gatekeepers: A National Survey of Judges on Judging Expert Evidence in a Post-*Daubert *World*, 25 LAW & HUM. BEHAV. 433 (2001)).

204. *Id.* at 1172.

205. See generally Sanders, *Paternalistic Justification*, *supra* note 37, for a thorough review of relevant literature.

206. Risinger, *supra* note 178.

lenges."[207] For example, Risinger notes that in the district court opinions he reviewed, civil defendants succeed in rejecting two-thirds of plaintiffs' evidence while prosecutors' challenged evidence was admitted in eleven out of twelve cases.[208]

IV. Implications of Empirical Research

A. Major Findings—Judges Have Adopted the Spirit, if Not the Text of *Daubert*

Taken together, these studies support two significant and somewhat paradoxical conclusions: *Daubert* has indeed raised the bar to admissibility, but judges are not frequently utilizing the reliability factors suggested in the decision. In other words, while judges are scrutinizing evidence more carefully, as required by *Daubert*, they appear to be developing their own criteria for determining admissibility.

First, *Daubert* raised the bar for admissibility in civil cases. According to both the RAND Study and the FJC Study, the frequency of pretrial hearings and motions for summary judgment has increased, and most importantly, judges are fulfilling their role as gatekeepers, excluding more evidence than before.[209] General changes in judicial practices and attitudes toward certain types of cases may have contributed to this change. Yet the evidence above suggests that the *Daubert* decision, whether directly or indirectly, has had a significant impact on these practices. As a result of *Daubert* challenges to admissibility, rates of exclusion increased significantly after 1993 in civil cases.[210] Judges began to scrutinize only physical-science type evidence more carefully.[211] This latter impact is what would be expected if *Daubert* was, at least partially, responsible for these changes.

Second, somewhat paradoxically, these studies suggest that the reliability factors provided in the decision are not well understood by many judges and, most importantly, are not frequently used to distinguish between admissible and inadmissible evidence. The relationship between these factors and the decision's effect of heightening the bar for admissibility is murky at best.

207. *Id.* at 99.

208. *Id.* at 109–10. Finally, it is necessary to acknowledge Pamela Jensen's Note, which concludes that the *Frye* and *Daubert* standards do not consistently lead to different outcomes in state courts for at least three types of evidence used in criminal cases. Pamela J. Jensen, Note, Frye *versus* Daubert: *Practically the Same?*, 87 MINN. L. REV. 1579 (2003).

209. RAND Study, *supra* note 3; FJC Study, *supra* note 3.

210. RAND Study, *supra* note 3, at 55.

211. *Id.* at 63.

According to the 1998 FJC survey, in cases where evidence was excluded, judges only pointed to problems with general acceptance, peer review, and insufficient theory testing less than 8% of the time, and dealt with falsifiability and error rates in less that 2% of cases (even though 18% of their exclusions were based on reliability considerations).[212] The FJC authors also report that the reasons for excluding testimony appear not to have changed significantly in the wake of *Daubert*.[213] Gatowski found that state judges, at least, do not have a clear understanding of at least two of the factors: falsifiability and error rates.[214] In the criminal context, the *Daubert* reliability criteria do not predict admissibility findings (with the exception of general acceptance, discussed below).[215]

This picture is somewhat complicated by the RAND finding that reliability generally and the reliability criteria specifically were discussed more frequently after the decision.[216] This trend turned around four years later.[217] After 1997, discussion of these factors decreased, with a corresponding increase in non-*Daubert* factors.[218] The second set of FJC surveys were distributed in 1998 and 1999 and hence may reflect this downturn. Even at their peak, each of these reliability criteria was discussed in the assessment of individual elements of evidence less than 40% of the time.[219] Additionally, data from the Groscup Study suggests that, at least in the criminal context, even when these criteria are addressed, they may not actually factor into the judge's decisions.[220]

One possible interpretation of the RAND data is that immediately following the decision, judges attempted to stay as close to the text as possible. As time went on, they grew emboldened to focus instead on factors that they felt were more pertinent, useful, or effective.[221] The RAND Study implies that parties likely responded to courts' heightened standards by submitting evidence that satisfied the reliability re-

212. FJC Study, *supra* note 3, at 323.

213. *Id.* at 330.

214. Gatowski, *supra* note 203. The significance of this finding—that judges do not grasp two of these factors—is called into question by Caudill & LaRue, *supra* note 82.

215. Groscup Study, *supra* note 3, at 353–56.

216. RAND Study, *supra* note 3, at 40.

217. *Id.*

218. *Id.* at xvi–xvii.

219. *Id.* at 39.

220. Groscup Study, *supra* note 3, at 353–56.

221. RAND Study, *supra* note 3, at 40–41.

quirement in some general sense, though not necessarily through the use of these factors.[222]

An alternative explanation is that judges applied the reliability criteria initially and parties responded by presenting evidence that complied with them.[223] Judges in subsequent years were then free to focus on additional criteria, as the basic factors were generally satisfied in most cases. This explanation is possible, but it strikes one as odd that courts would essentially stop discussing the reliability factors simply because most evidence conformed to them. Moreover, given state judges' apparent confusion over the meaning of at least two of these factors,[224] it is more likely that courts increasingly abandoned the factors, instead developing their own criteria for reliability. It bears mentioning that even if this latter explanation is correct—even if reliability factors have been addressed less frequently since 1997 because parties have started submitting evidence that meets these standards—the fact remains that judges in later years were not using the reliability criteria to distinguish between admissible and inadmissible data.

To summarize, the data suggests that *Daubert* had a rather profound effect on the admissibility of evidence. The relationship between the reliability criteria and this effect, however, is unclear. What seems to be missing is an explanation for how, mechanically, *Daubert* has effectuated this change. At this junction, the Cheng & Yoon Study provides critical insight.

B. Analysis and Discussion of Cheng & Yoon Study Findings

This study strongly implies that what is important about *Daubert* is *not* the doctrinal test included in the decision, but rather the awareness that the decision generated. Most of the changes effected by *Daubert* appear to be the result of simply informing judges that they should assume the role of gatekeeper. By focusing closely on the text of the decision, prior critics have given too little attention to the bigger picture—*Daubert*'s place in a cultural shift toward greater skepticism about scientific evidence.

By Cheng & Yoon's account, judges in jurisdictions that adopted *Daubert* do not appear to be treating evidence differently from judges in *Frye* jurisdictions, at least not in any manner that actually matters to

222. *Id.* at 41.
223. *Id.* at 62.
224. Vidmar & Diamond, *supra* note 203, 1172–73.

defendants.[225] What does this mean? For one thing, it reinforces the idea that the reliability criteria currently play an insignificant role in how judges distinguish between admissible and inadmissible evidence since the criteria are not included in the *Frye* test. But its implications are even broader, for this study suggests that the actual text of *Daubert* is far less important than the decision's cultural and educative impacts. To the extent the decision had a real effect on admissibility, it did so primarily by informing judges that they should function as gatekeepers to ensure that bad science does not make its way into the courtroom. If Cheng & Yoon are correct, it would seem that judges in all jurisdictions have adopted this role and that the *Daubert* decision is largely irrelevant in explaining what is happening in the courtroom.

The methodology of this study is novel and subject to serious criticism. Nonetheless, this interpretation of *Daubert*'s impacts is consistent with the paradoxical fact that the decision, which in some sense was meant to "liberalize" the evidence rules, in fact tightened them. As discussed above, the Court decided *Daubert* at a time when the problem of junk science was widely debated. *Daubert*'s impacts would seem more consonant with a cultural phenomenon of heightened skepticism toward science, than with the decision's liberal thrust and literal message that scientific theories need not be generally accepted to be admissible. Moreover, this interpretation is consistent with the fact that judges tend to rely on non-*Daubert* factors in rendering determinations as to admissibility.[226]

225. Cheng & Yoon Study, *supra* note 4, at 503.

226. This data also tends to favor the interpretation that the surprisingly small amount of judicial attention to the reliability factors in recent years is not explained by the theory that evidence generally satisfies these factors. Presumably, if judges in *Daubert* jurisdictions are still looking at these criteria, we would probably expect to see some differences in removal rates in those jurisdictions that have stuck to the *Frye* standard versus those that have adopted *Daubert* because attorneys who believe that proffered evidence does not satisfy these criteria would be more likely to request removal.

This discussion contains one major caveat: while many of the criticisms discussed specifically addresses toxic tort cases, the studies above that reviewed civil cases included any cases containing expert testimony or challenges to expert testimony, with the exception of the Cheng & Yoon Study, which focused on torts generally. Toxic tort cases are a unique subset of these, and it is not possible to say whether the generalizations made here necessarily apply in full to this subset. The Groscup Study, while focusing on criminal cases, did find lengthier discussions of error rate and general acceptance for cases involving scientific experts than for those involving other types of experts. Groscup Study, *supra* note 3, at 369. Thus, it is possible that the reliability factors are being applied at a far greater rate in the context of scientific evidence, and possibly even more so in the toxic tort setting. Then again, increased discussion of factors does not necessarily imply that the influence of these factors has increased. Future research focused specifically on scientific evidence and toxic tort cases would be particularly informative.

None of these studies was designed to directly address the question of whether judges are excluding too much evidence or the wrong evidence, or whether the bar for admissibility has been set too high, or too low for that matter (specifically in civil cases).[227] One possible exception to this is Risinger's observation that criminal defendants almost always lose their admissibility challenges while civil defendants usually prevail.[228] The disparities noted by Risinger are rather striking and suggest that something different is at work in the criminal context. Judges may exhibit a bias in favor of evidence proposed by prosecutors. It is also possible, however, that prosecutors submit better, more reliable evidence than civil plaintiffs or that court appointed defense counsel, with heavy caseloads and fewer resources than civil defense attorneys, launch less persuasive challenges to admissibility. More research is needed to determine which of these factors are responsible for Risinger's findings.[229] Considering that this evidence only includes relative admissibility rates, it is impossible to determine whether the standard for admissibility in civil cases is too strict, the standard for admissibility in criminal cases is too lax, or both.

C. Do Current Criticisms of *Daubert* Miss the Real Issue?

Most scholars who have contributed to the academic dialogue over the merits of the *Daubert* decision have done so without the benefit of this recently accrued empirical data. The observations discussed in the preceding section shed considerable light on the ways in which *Daubert* may or may not have fallen short of expectations about what such an admissibility standard should provide.[230]

227. Identifying the mechanism by which *Daubert* has caused an increase in excluded evidence does not necessarily equate to finding the mechanisms by which the application of *Daubert* may exclude evidence that otherwise should be admitted. However, by discovering what factors contribute most to a judge's admissibility finding in general, we get one step closer to identifying which factors may be related to poor admissibility determinations.

228. Risinger, *supra* note 178, at 99.

229. It bears mentioning, however, that courts' frequent willingness to admit somewhat dubious evidence, including handwriting and hair samples, suggests that a difference in the quality of evidence submitted by civil plaintiffs versus prosecutors is probably an incomplete explanation. *Id.* at 105–12.

230. Unfortunately, many of the criticisms addressed in Part II either do not lend themselves to empirical testing or were not addressed by this data. The second half of this section makes suggestions for future research to explore those criticisms that can be tested. Although many of the criticisms concerning *Daubert* addressed in Part II could not be tested empirically, some are at least consistent with these findings. For example, while it is obviously impossible to empirically test whether *Daubert* violates the spirit of the Seventh Amendment, the evidence above at least supports the notion that juries are hearing a considerably smaller proportion of proffered evidence. Furthermore, while I have not

First, the empirical observations above suggest that the reliability criteria probably are not particularly important in how judges make admissibility determinations. On the one hand, this would seem to render futile inquiries into whether or not *Daubert* incorporated a flawed concept of good science by choosing poor indicators of reliability (for example, over-reliance on statistical significance or peer review) and whether or not judges are misapplying the criteria listed in the decision (except insofar as they fail to apply them!). While it is not possible to test whether the decision embraces a faulty philosophy of science, we could conclude that even if it does, this fault probably has little impact on admissibility decisions. As the Cheng & Yoon Study implies, perhaps we are wasting time debating the merits of the doctrinal test in *Daubert*.

This response, however, may be incomplete because it begs the question of *why* judges do not find the *Daubert* factors useful.[231] We have seen at least two possible answers to this question. The reliability factors simply may not be applicable to much of the evidence that comes before judges. In other words, perhaps *Daubert* reflects an idealized version of science that is largely unhelpful given the realities of scientific evidence.[232] Or the explanation may be even simpler: perhaps judges do not have a sufficient understanding of what these factors mean or how to apply them. Certainly the Gatowski study supports this conclusion, at least among state court judges.[233] Either way, the data would seem to imply that *Daubert* has failed to provide a framework that judges, at least in recent years, find useful in distin-

thoroughly summarized the extensive evidence concerning the relative proficiencies of judges and juries in analyzing expert testimony, this data does raise concerns that paternalistic justifications for the gatekeeping role may be misguided. *See, e.g.*, Sanders, *Legal Perceptions*, *supra* note 9. Other criticisms that do not readily lend themselves to empirical testing are whether plaintiffs have suffered a psychological loss from being denied the opportunity to air their cases before a jury, whether defendants have been inadequately deterred because less information has been brought to the public's attention, the criticism that judge-made law under *Daubert* violates the principles of *Erie*, the criticisms that *Daubert* should have incorporated a probabilistic notion of causation for reasons of improving justice for plaintiffs, and finally, the argument that this decision exacerbated the cultural conflict between science and the law.

231. As discussed above, though it may be the case that the reliability factors have helped improve the quality of evidence presented, the reliability factors do not appear to be particularly useful in recent years for helping judges distinguish between good and bad evidence.

232. *See generally* Caudill & LaRue, *supra* note 82.

233. *See* Vidmar & Diamond, *supra* note 203, at 1172–73 (discussing Gatowski, *supra* note 203).

guishing between good and bad science, or admissible and inadmissible evidence.[234]

These findings also illuminate concerns that judges may not be competent to perform the task of distinguishing between good and bad science.[235] Because the data above does not provide a complete picture of how, exactly, judges make admissibility determinations, it is impossible to reach any definitive response to this concern. But the fact that judges are not making any great use of the reliability factors in rendering decisions helps inform our concerns about their competency. By straying from the reliability factors, judges may be applying inappropriate criteria and rendering inconsistent judgments.[236] For example, scholars have taken issue with some judges' interpretation of *Daubert* as requiring epidemiological data or specifically epidemiological data that demonstrates a doubling of risk.[237]

On the flip side, judges may be exploiting the elasticity of the decision in a positive way. They may be correcting for *Daubert*'s failure to account for variation in the type of data produced by different scientific fields. In other words, judges may be adjusting to the realities of what scientific evidence looks like in the world. Some scholars appear to be encouraged by a judicial willingness to adopt a pragmatic approach towards scientific evidence.[238] As indicated above, the RAND Study found that judges, once familiar with their role as gatekeepers, developed their own factors for determining the admissibility of evidence.[239] Perhaps those scholars who take issue with the inclusion of peer review or statistical significance in the *Daubert* test should be encouraged by judges' flexibility in analyzing evidence. At the very least, these findings seem to appease fears that judges are simply applying the reliability criteria in a formulaic and unthinking manner.

234. It is important to remember that the *Daubert* Court explicitly disclaimed providing some sort of checklist for reliability. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 593 (1993). Nonetheless, these factors were intended to provide judges with some guidance and it would appear that such guidance has been largely abandoned. *See id.*

235. For a discussion on what sort of epistemological inquiries judges are competent to undertake, see Brian Leiter, *The Epistemology of Admissibility: Why Even Good Philosophy of Science Would Not Make for Good Philosophy of Evidence*, 1997 BYU L. Rev. 803, 805, 814–19 (1997). This criticism is related to arguments concerning the philosophy of science adopted by the decision.

236. This hypothesis is supported by the observation that courts have come to inconsistent conclusions regarding the effects of the drug Parlodel. TELLUS PUBLICATION, *supra* note 36, at 11–12.

237. *Judicial Boundary Drawing*, *supra* note 42, at 39–41.

238. *See generally* Caudill & LaRue, *supra* note 82. Judges may also be taking into account cultural and rhetorical elements of science. *See id.*

239. *See generally* RAND Study, *supra* note 3, at 40.

Despite its failure to provide useful criteria for judges, *Daubert* appears to have been successful in elevating judges to the position of gatekeeper. Judges appear to be scrutinizing evidence more carefully and admitting less of it. If Cheng & Yoon's hypothesis is correct, *Daubert* and the debate surrounding the decision had this effect not only in jurisdictions in which the *Daubert* standard has been adopted, but even in jurisdictions that have not adopted it.

Indeed, the evidence would seem to suggest that the decision reflects, or perhaps has spawned, a cultural phenomenon in the law whose primary characteristics are skepticism toward scientific evidence and a more active role for the judge in assessing evidence. Once again, this development could explain how a decision that was supposed to liberalize the admissibility of evidence by allowing in methodologically sound minority opinions,[240] which had been previously excluded under *Frye*,[241] has, in fact, had the opposite effect. After *Daubert*, a lack of general acceptance was a better predictor for a finding of inadmissibility than before *Daubert*.[242] This finding is not only surprising but also ironic; at the same time that judges are taking a significantly more active role in assessing the quality of evidence on their own terms, they are demonstrating increased deference to the majority view of the scientific community. Most importantly, this finding suggests that *Daubert* is being misapplied to some degree. A faithful application of the decision should not result in an increased role for general acceptance.

Overall, there is much explanatory power in Cheng & Yoon's theory that the effects from *Daubert* have more to do with increasing awareness about bad science and with a cultural shift towards greater skepticism about scientific evidence than with the actual doctrine of the case. As seen through this lens, the decision either accomplished a tremendous amount or accomplished very little. While it helped revolutionize how judges across the country in all jurisdictions perceive scientific evidence and their role in assessing it, it also provided a standard with little usable content. True, the decision undoubtedly did do something important—something worth talking about. But these studies also suggest that if we are going to turn a critical eye to the *Daubert* decision, we should be faulting it not for what it did, but for what it has failed to do. Or perhaps, we should refocus many of our criticisms away from the decision itself and onto the development of a

240. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 588–89 (1993).
241. *See* Sanders, *Legal Perceptions*, *supra* note 9, at 140.
242. Though this factor is still only mentioned in a small percentage of cases.

culture in the law that treats science with a certain type of scrutiny and skepticism. To the extent that we want to improve judicial practices with regard to admissibility, we may want to either incorporate into the standard those factors that judges actually find useful in assessing evidence, or to focus more on "soft" solutions that "increase the judiciary's understanding of scientific concepts and processes."[243]

D. Suggestions for Future Research

The studies above raise serious questions as to what factors judges are using to distinguish admissible from inadmissible evidence and why the reliability factors enumerated in *Daubert* are not more regularly utilized. It would be particularly interesting to ascertain why judges are applying the *Daubert* factors so infrequently and why these factors seem not to correlate well with admissibility decisions (except for general acceptance). By simply asking judges, via a survey instrument, what factors they are utilizing and why, we could significantly improve our understanding of this phenomenon.

None of these studies provides sufficient information on what other reliability-indicating factors judges are using, nor on whether the decision should have contained other reliability factors. Although the RAND Study coded for many factors that entered into reliability assessments, those factors that were included in the published findings were relatively uncontroversial. One potential line of future research would involve identifying a list of reliability factors (including controversial ones) and then conducting a review of decisions similar to the RAND Study (particularly focusing on toxic tort cases).

The criticisms that *Daubert* and *Joiner* encourage sequential review of evidence for admissibility and may confuse admissibility and sufficiency standards[244] are consistent with the general finding that less evidence is being admitted.[245] Nevertheless, the studies conducted so far have not captured data that indicates whether judges are, indeed, evaluating evidence in this manner. In the future, researchers may want to ask several questions when looking at admissibility decisions, and perhaps also at the transcripts of *Daubert* hearings: Does it appear that judges are considering the relevance and reliability of evidence

243. Cheng & Yoon Study, *supra* note 4, at 504. Curiously, judges may not even be fully aware of the impact *Daubert* has had on their practices, as indicated by the FJC Study finding that a majority of judges failed to report actual changes in their management of cases after *Daubert*. FJC Study, *supra* note 3, at 329.

244. *Intellectual Due Process*, *supra* note 38, 1075–76; Capra, *supra* note 45, at 754–55.

245. *See generally*, RAND Study, *supra* note 3.

sequentially or are they evaluating it holistically? Do judges appear to be making sufficiency judgments when they should be focusing on reliability or are they keeping these two inquiries distinct? Finally, do differences in these procedures result in different outcomes?

Another fertile topic for future inquiry is whether *Daubert* has reduced the frequency with which defendants produce evidence. As mentioned earlier, this criticism is particularly relevant in the toxic tort context, where encouraging defendants to produce scientific evidence may help improve our position of relative ignorance about chemical exposure. At first glance, the FJC Study would seem to refute this finding. Its authors observed that the percentage of trials containing experts testifying for plaintiffs and the percentage containing experts testifying for defendants remained relatively consistent from 1991 to 1998.[246] However, this data includes only cases that went to trial.[247] The rationale behind the idea that defendants would produce less evidence is that early in the case they may be able to defeat the plaintiff by criticizing her evidence, winning a motion for exclusion, and prevailing in a summary judgment motion all without having to proffer any evidence. Once a case is destined for trial, the defendant would be far more likely to present her best evidence. Research directed at answering this question could address the question of whether, for all cases that at least reach summary judgment, defendants are providing less scientific evidence than they used to.

Finally, as suggested by the authors of the RAND Study, research could be conducted to compare judges' actual decisions on admissibility with independent experts' assessments. This could produce very interesting and valuable results. However, were such a study to be carried out, it would be important to remember that not all scholars agree with the notion that the best scientific judgment of good evidence necessarily equates to the best legal judgment of good evidence.

Conclusion

This Article has provided an overview of criticisms of the Supreme Court's decision in *Daubert* and a summary of recent empirical data showing how *Daubert* is being applied by judges and how parties to litigations have responded to it. Reviewing this data, the Article concludes that *Daubert* has raised the bar for admissibility, and, at least

246. FJC Study, *supra* note 3, at 318–19.
247. *Id.* at 318.

in recent years, that the reliability factors enumerated in the decision are not widely used and probably are not the mechanism by which *Daubert* has affected admissibility.

Criticisms of *Daubert* should focus more on the decision's failure to establish a useful list of criteria that, in practice, assists judges in making admissibility determinations. This Article adopts Cheng & Yoon's hypothesis that *Daubert*'s most important contribution to the development of evidentiary standards has more to do with *Daubert*'s educative function than with its doctrinal text. The decision has created or contributed to a large cultural shift while failing to provide useful content that would ensure that this transformation is effectuated in a consistent fashion. This cultural shift is likely responsible for increasing, rather than decreasing, the relative importance of general acceptance as a factor in making admissibility determinations. This result is inconsistent with the thrust of *Daubert*.

Finally, this Article has made recommendations for future research to help answer the many intriguing questions left open by prior studies. Chief among these is research aimed at determining what criteria judges are actually using to distinguish between admissible and inadmissible evidence. This is the key question to which scholars must now turn a critical eye.